AI-EVAL

Evaluation, Observability and Guardrails

Without a golden dataset and release gates, every AI release is an experiment on your users.

DURATION	GROUP	FORMAT
2 days	**up to 12 people**	In-house · Remote · In the lab · German or English

THIS EXISTS AFTERWARDS

- A golden dataset for your own system: curated cases with expected results, edge cases and deliberately hard queries
- Implemented quality metrics for groundedness, relevance and completeness, including a calibrated LLM-as-judge with human spot-check validation
- Tracing across the full chain from request through retrieval and model calls to the answer, with cost and latency per step
- A cost and latency budget per request type with alert thresholds
- Defined release gates: which measurements a deployment must pass and when a rollback is triggered

AGENDA

Day 1	The blind flight diagnosis · Building the golden dataset · Implementing metrics: groundedness, relevance, completeness · The first evaluation run · Day result: a reliable baseline
Day 2	Tracing: making the chain visible · Cost and latency budget · Guardrails before and after the model · Release gates and regression detection · Day result: the gate holds

YOUR TRAINER

Dino Bordonaro

Microsoft MVP for eight consecutive years (Microsoft Azure · Cloud and Datacenter Management), Microsoft Certified Trainer, almost 30 years in IT. Builds sovereign AI systems on own hardware (Azure Local, Foundry Local, NVIDIA GPUs) and trains on exactly that. More than 800 professionals trained, every concept proven in his own enterprise lab in a TÜV-audited datacenter.

€15,800 flat, plus VAT

This track enables your team to build and run its own measurement system. It is neither an audit of your production system nor a managed monitoring service, for those we talk about a project.

AI-EVAL

Dino Bordonaro · BORDONARO IT GmbH & Co. KG · info@bordonaro-it.de · +49 6232 298 53 - 0 · bordonaro.ai